DOCUMENT RESUME

ED 480 044 CG 032 617

AUTHOR	Lundberg, David; Kirk, Wyatt
TITLE	A Test User's Guide to Serving a Multicultural Community.
PUB DATE	2003-08-00
NOTE	11p.; In: Measuring Up: Assessment Issues for Teachers, Counselors, and Administrators; see CG 032 608.
PUB TYPE	Information Analyses (070)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Cultural Influences; Educational Assessment; *Educational Testing; Evaluation Methods; *Test Bias

ABSTRACT

Testing is one means of viewing differences among individuals. Culture is another means. When we mix testing and culture together, the results are fascinating and often confusing. Generally, we test individuals in an attempt either to serve them or reward them, and if we want to reward people, there is a strong desire and need to be fair. However, fairness is not easy to define or implement in the volatile arena of testing and culture. This chapter explores various recommended actions and strategies for pursuing fairness in the testing process. (Contains 10 references.) (GCP)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 480 044

A Test User's Guide to Serving a Multicultural Community

By
David Lundberg
Wyatt Kirk

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- ☐ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CG032617

BEST COPY AVAILABLE

Chapter 9
A Test User's Guide to Serving a Multicultural Community

David Lundberg & Wyatt Kirk

Testing is one means of viewing differences among individuals. Culture is another means. When we mix testing and culture together, the results are fascinating and often confusing. Generally, we test individuals in an attempt either to serve them or to reward them, and if we want to reward people, there is a strong desire and need to be fair. However, fairness is not easy to define or implement in the volatile arena of testing and culture.

One way to pursue fairness in testing is to assess students in a standardized manner, using the same methods, content, administration, scoring, and interpretation for everyone. A major problem with this "equality" approach is that if certain groups differ on irrelevant knowledge or skills that affect their ultimate performance on the test, then bias exists (Lam, 2001). The question arises, "Can identical assessment really be fair to different cultural groups?"

Another way to pursue fairness is to tailor the testing process to each individual's special background (i.e., his or her culture). The major problem with this approach is in ensuring that the results of different testing processes are truly comparable across groups (Lam, 2001). Differing assessments may seem more equitable, but are they really more fair? This is the dilemma of the test administrator or user who serves a multicultural community.

Culture and Assessment

What is culture really? When we view and define culture broadly, the factors involved seem almost endless. Age, sex, place of residence, social status, educational level, income, nationality, ethnicity, language, religion, and a host of affiliations from family of origin to social cliques to professional grouping are all variables in the broad definition of culture (Pedersen, 1991). When we define culture more narrowly, with respect to just a few variables, then group people according to those

variables, differences become noticeable. For example, if we compare 14-year-old White females to 14-year-old Hispanic males, some common characteristics will obviously differ between the two groups. In the best sense, making generalizations and intelligent judgments about these cultural differences can provide a background for understanding each person's uniqueness. When judgments about groups become rigid, however, and that picture of a unique human being is lost, stereotyping and its negative effects creep in (Sue & Sue, 1990).

Another means of comparison is testing, and comparisons seen through test results can be valuable. Some tests are interpreted in either a bipolar or a neutral manner, meaning that any individual score is not considered better or worse than any other score. Personality tests and interest inventories given by counselors fall into this category. Examples of these neutral or bipolar tests are the Myers-Briggs Type Indicator (MBTI) and the Strong Interest Inventory (SII). The MBTI is a test that assesses individual personality on four bipolar scales. Whether an individual's score on the first MBTI scale is more Introverted or more Extroverted is considered as neither superior nor inferior. It simply forms a basis for comparison. Likewise, the SII gauges a person's interest in a wide range of occupational areas. Whether a person expresses strong interest or little interest in any particular vocational area is, again, of no inherent value positively or negatively, but it can be valuable as a comparison to that person's interest in the other occupational areas.

In contrast, many educational tests *are* high-low in their interpretation. This high-low orientation generally results in a benefit for the best scoring individuals. They often receive higher grades or better treatment as a result of testing.

Within school systems, most tests are produced locally by teachers who seek to measure the achievement or learning of their students. It is assumed that each student in the teacher's class was exposed to the same instruction and that the test is the same for each student. The teacher compares individual scores on locally produced tests to evaluate the progress of the various students. These locally produced tests are obviously high-low in their orientation.

Standardized tests are generally developed by large companies and often distributed nationally. They are used with broad audiences and given with the assumption that testing conditions and the test itself are the same for all students. The purpose of standardized tests is to compare the scores of a single student or a group of students to the scores of a national sample of students or to a chosen reference score.

Just like the locally produced tests, standardized tests are generally high-low in their purpose and interpretation.

The assumption of sameness for any test, whether locally produced or standardized, is problematic. Although an identical test can be given to different students, no two students are identical. Therefore, the test can never be the same for all individuals. The problem with testing is not that we don't have standardized students, however. Tests are meant to discriminate among individuals. The problem is that we don't have standardized cultures, so differences in culture interfere with the simple comparison of students. Tests may be somewhat similar for people of similar culture, but those tests can be markedly different for people of different cultural groups.

Standardized tests are developed and normed using a particular sample, and historically in our society that sample has been predominantly white and middle class. Today, many test publishers make an effort to include students of all types in their test development process so that the norm group is representative of the target population. When this is not feasible, efforts are made to "prove" that standardized tests are suitable for groups that were not represented or were little represented in the original test development and norming process. In either case, every test user should carefully screen the technical background information of any test to determine its applicability to people of color. Large amounts of time and resources are expended developing efficient, relatively short tests with questions that result in a predictable pattern of correct responses. But there has always been, and there continues to be, great controversy over applicability of standardized tests to all cultures.

Recommended Actions and Strategies

The purpose and use of the comparative results of tests are the real issues in all testing, but particularly in standardized testing. The burning question is, "What comparisons are being made and for what purpose?" Tests are best used when they serve the test taker. The test user should look upon a test as a tool to further the development of the person being assessed. It is very common to see the results of standardized tests being used to categorize individuals rather than to serve them. In addition, the results of standardized tests are now being extended to categorize schools and school systems.

We live a world of incredible diversity, limited resources, and strong desires for quick, efficient answers. Given the variety that exists

among human beings and the desire to compare individuals by using tests, how can test users better utilize those tests for the benefit of the various populations they serve? There are a number of crucial multicultural factors in testing. Understanding these factors is the first step in using tests constructively to help diverse populations.

Differences in Communication and Learning Styles

No one prescribed method or model of teaching or learning fits all people. Many teachers and counselors use vary their styles of communication and instruction in an effort to evoke the best results from their students and clients. Skillfully alternating and integrating teaching styles allow material to be presented in several ways with the hope that one of the styles may engage the student in the learning process. Additionally, there is great benefit in not boring students with the same repetitive method.

Just as people think and learn differently from each other, we need to assess their resulting competencies in various ways. Too often we are tempted to assume automatically that a person with a lower score on a given test is less advanced in general than his or her counterpart who achieved a higher score. What we know for sure in such a situation is that the higher scoring student has succeeded in answering the particular questions on that particular test in the particular way they were communicated. If test content is well aligned with curriculum standards, this is also an indication that the higher scoring individuals are more closely approximating those standards. However, the generalization that the lower scoring individuals are less advanced is often fallacious.

We just don't know enough about the learning styles prevalent in many cultural groups and how those learning styles are best assessed. There has been far too little research in these areas. We tend to use communication patterns and teaching methods developed over many years that basically work with the majority population. We implicitly expect minorities to adapt to the majority style. If they do, they are competitive. If they don't, they are low performing. If certain minority members excel, we tend to think of them as superior, but we lose sight of the fact that they are also operating extremely effectively outside their normal culture, a skill that majority members are seldom asked to develop.

Long-Term Poverty

There is a somewhat hidden minority in America. This group contains Whites, Blacks, Hispanics, Native Americans, and many other subgroups. It is spread across all geographical areas, and it is both urban and rural. This minority is the long-term poor. There are disproportionate percentages of Blacks, Hispanics, and Native Americans in this group, which seriously distorts an examination of group performance in testing.

When studies of low test performance by minorities are corrected by statistically controlling for the effect of socioeconomic status (SES), the low performance is just as evident (and often more evident) with those who are poor as it is for minorities. In other words, the primary issue is often one of income, not more visible factors like race or ethnicity (Abbott & Joireman, 2001; Betts, Reuben, & Danenberg, 2000).

Unfortunately, it is much easier to correlate low scores with those more visible factors, and this is constantly done. We continually read that Black students or Hispanic students or Native American students score differently (usually lower) on tests than the majority group. Students do not walk around with signs proclaiming their gross household income, and however silly that statement appears, household income is often a more accurate predictor of test scores than are ethnicity or race (Dixon-Floyd & Johnson, 1997; Fergusson, Lloyd, & Horwood, 1991). Test users should favor tests that are developed or normed with consideration specifically for low-income students.

Expectations, Confidence, and Motivation

Because of the long-term conditions of poverty, many people of color suffer from chronically low expectations, confidence, and motivation. These problems cannot be overemphasized, and they certainly have no quick, effective solutions. Many members of minority groups wage lifelong battles to overcome these limitations. In our society, low SES corresponds with fewer resources for schools, less qualified teachers, and fewer advanced course offerings (Betts, et al., 2000).

Viewing each test taker as an important individual with a unique combination of characteristics and undeveloped potential should be the first step in any test user's approach. The characteristics vary among students, and a student's potential may lie in surprising areas, but seeing that person's uniqueness can be the first crucial step in providing expectations, confidence, and motivation to any student who doesn't fit the mold.

Differing Dialects

In the United States, we tend to think of dialects as something found in Europe or among tribes in third world countries, but differing languages are a reality in this country. This reality goes beyond varying communication styles, and it goes beyond having a different mother tongue. In many inner city environments, for example, the English words and phrases minority members use to communicate on a day-to-day basis nearly comprise a different language.

When students from these other cultures, such as inner city children, take standardized tests that are written in the language of middle- and upper-class students, those children are reading a somewhat foreign language. The resultant test scores are almost always lower than those of the majority.

Test Readiness and Hidden Talent

Few people love tests. But as a matter of survival and advancement, many learn how to prepare for and take tests, and they view testing as important for their future. Many students from low income brackets are not socialized to view tests as important. Other factors are more crucial to their success in school or the everyday world than getting a good grade on a test. Social prowess, leadership, nonverbal communication, and a host of other factors may be more important to many members of minority groups. It is incumbent upon test users and administrators to communicate effectively the importance of testing in today's society. The need for equitable access to test preparation programs should be continually stressed.

Tests don't do a very good job of evaluating creativity or imagination. They have difficulty measuring entrepreneurial drive or initiative. There aren't any tests that are very effective at assessing imagery, the ability to visualize a solution to a problem. Tests are good at demonstrating which students are able to take in, hold, and repeat information presented in certain ways. Tests are good at rewarding certain cognitive processes.

Speed in answering is a prime factor in scoring well on tests. Most tests favor those students who are skilled at memorization and can respond rapidly to the specific test format. A lack of tests that adequately identify important skills along with a lack of test readiness among youth of color (Castenell & Castenell, 1988) limit identification of certain talented individuals.

Other Forms of Assessment

Most standardized testing is in a multiple-choice, matching, or true-false format. There are some advantages to these formats in terms of flexibility in addressing broad areas of content and in measuring specific, sometimes very complex, thinking processes. An overwhelming advantage of the multiple-choice format is that it is inexpensive to score.

Other forms of assessment add more information and a broader picture in evaluating individual performance (Supovitz, 1997). Examples of these alternative instruments are essay questions and performance assessments. These forms are more expensive, and they are prone to criticisms of subjectivity. Individual evaluators have considerable leeway in grading performance when looking at an essay or performance assessment. Biased evaluations or favoritism can be problems; however, standardized multiple-choice tests have inherent bias and favoritism because of the factors mentioned previously. Research with alternative assessment modes has indicated some potential to decrease inequities seen with standardized tests. However, care must be taken in the development of these assessments (Supovitz & Brennan, 1997).

Conclusion

As test users, recognizing that we live in an imperfect world does little to help the individual student who stands before us looking for education and training that will equip him or her for a successful life. Our challenge is immediate, society changes very slowly, and that young man or woman is maturing rapidly.

Our first step is to recognize each individual as a person of great value and undeveloped, unknown talents. No single test or battery of tests of similar format can ever explain a person. No test can level the field or compensate for all the diversity present in a single school, much less in our society. And no evaluation instrument can replace the importance of one human being interacting with another.

Tests provide us with information, not answers. They provide the substance of conversation, not decisions. Answers and decisions about people or groups of people are not what education is about. Our educational system should produce motivated, capable, and confident graduates who are able to satisfy themselves and contribute to our world.

The encouragement and intelligent explanations a test user provides to a test taker form the basis for that student's personal

development long after the results of all tests are forgotten. No test can stand alone. Use assessment that is based upon multiple tests with multiple formats. Use other forms of assessment that are realistic, even if they are more labor intensive. If you use standardized tests, choose those that have been developed and normed with full consideration for low-income and minority students. Explore or develop tests that are suitable for members of minority groups, and invite test makers to develop standardized tests that are specific to minority cultures. Don't elevate the results of any one assessment to a supreme degree. Use tests to serve the test taker. Never allow the student to become a servant to the test. In the end, your support of the test taker can be the most important element in assessment, and that support can produce a lasting effect in a student's life.

References

Abbott, M. L., & Joireman, J. (2001). *The relationships among achievement, low income, and ethnicity across six groups of Washington state students.* (Report No. WSRC-TR-1). Lynnwood, WA: Washington School Research Center. (ERIC Document Reproduction Service No. ED454346)

Betts, J. R., Reuben, K. S., & Danenberg, A. (2000). *Equal resources, equal outcomes? The distribution of school resources and student achievement in California.* San Francisco, CA: Public Policy Institute of California. (ERIC Document Reproduction Service No. ED 451291)

Castenell, L. A., Jr., & Castenell, M. E. (1988). Norm-referenced testing and low-income Blacks. *Journal of Counseling and Development, 67,* 205–206.

Dixon-Floyd, I., & Johnson, S. W. (1997). Variables associated with assigning students to behavioral classrooms. *Journal of Educational Research, 91*(2), 123–126.

Fergusson, D. M., Lloyd, M., & Horwood, L. J. (1991). Family ethnicity, social background and scholastic achievement: An eleven year longitudinal study. *New Zealand Journal of Educational Studies, 26*(1), 49–63.

Lam, T. C. M. (2001). Fairness in performance assessment. In G. R. Walz & J. C. Bleuer (Eds.), *Assessment: Issues and challenges for the millennium.* Greensboro, NC: ERIC-CASS.

Pedersen, P. B. (1991). Multiculturalism as a generic approach to counseling. *Journal of Counseling and Development, 70,* 6–12.

Sue, D. W., & Sue, D. (1990). *Counseling the culturally different.* New York: John Wiley and Sons.

Supovitz, J. A. (1997, November 5). From multiple choice to multiple choices: A diverse society deserves a more diverse assessment system. *Education Week,* p. 34.

Supovitz, J. A., & Brennan, R. T. (1997). Mirror, mirror on the wall, which is the fairest test of all? An examination of the equitability of portfolio assessment relative to standardized tests. *Harvard Educational Review, 67*(3), 472–506.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[ ] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[X] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").